AD _____

Award Number: W81XWH-10-1-0221

TITLE: Use of Buccal DNA Methylation Profiles to Risk Stratify Current and Former Smokers for Lung Cancer

PRINCIPAL INVESTIGATOR: Jason W. Chien, M.D., M.S.

CONTRACTING ORGANIZATION: Fred Hutchinson Cancer Research Center
Seattle, WA 98109-1024

REPORT DATE: July July 2011

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT:

☒ Approved for public release; distribution unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
31/jul/2011	Final	01 JUL 2010 - 30 JUN 2011

4. TITLE AND SUBTITLE

Use of buccal DNA methylation profiles to risk stratify current and former smokers for lung cancer.

5a. CONTRACT NUMBER
W81XWH-10-1-0221

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Jason W. Chien, M.D., M.S.

jchien@fhcrc.org

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Fred Hutchinson Cancer
Research Center

Seattle, WA 98109

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research
And Materiel Command
Fort Detrick, MD 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The main goal of this proposal was to focus on the "identification and development of tools for screening or early detection of lung cancer" by evaluating the methylation profile of DNA from buccal cells as a risk stratifying tool for curent and former smokers at risk for developing lung cancer. Our initial findings indicate that 1) the methylation patterns of the blood and buccal DNA appeared to be distinct, 2) the epigenetic profile of buccal and blood DNA is most different among current smokers, suggesting that tobacco smoke exposure affects buccal DNA differently from blood DNA, and 3) smoking exposure is associated with more hypermethylated loci in buccal DNA when compared to former and never smokers. These data will direct our approach to evaluating a case control study of lung cancer cases and noncases, which have already been genotyped. These results provide strong evidence supporting the possibility that buccal epigenetic profiles can be informative in identifying individuals at high risk for lung cancer.

15. SUBJECT TERMS

lung cancer, early detection, buccal DNA, epigenetics, methylation

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			USAMRMC
U	U	U	UU	16	**19b. TELEPHONE NUMBER** *(include area code)*

Table of Contents

Introduction

The main goal of this proposal was to focus on the "identification and development of tools for screening or early detection of lung cancer" by exploring an innovative new concept in lung cancer screening that uses the methylation profile of DNA from buccal cells to risk stratify current and former smokers at risk for developing lung cancer. This was to be achieved through two Specific Aims:

Aim 1: Identify a panel of genes whose methylation profile in buccal cells can reliably differentiate between lung cancer cases and non-cases.

Aim 2: Evaluate whether the methylation profile identified in Aim 1 can be used to forecast the development of lung cancer within one year after the buccal specimen was obtained.

Body

Upon receipt of the award, our first priority was to obtain all of the necessary biospecimens from the large prospective cancer epidemiology study specified in the proposal. During the wait for the specimens, we conducted multiple proof-of-concept experiments in the laboratory, demonstrating that the methylation array performed well on buccal specimens collected from our lung cancer clinic, yielding accurate and reproducible data.

Upon evaluation of the specimens for the cases and controls, it became apparent that the quality of the specimens was suspect, and there was potential for contamination of the specimens. Further review of the collection protocol with the principal investigator of the epidemiology study revealed that not all of the study participants followed the specified instructions for sample collection.

Based upon this, we designed several experiments to test alternative hypotheses that would contribute to this work. First, we evaluated whether buccal specimens collected using different protocols that varied the time from collection to freezing may have affected the methylation status of the DNA. Evaluation of data from this experiment revealed that this did not alter significantly the methylation status of the DNA specimens.

Second, we performed a comprehensive analysis of buccal methylation profiles by comparing them with blood methylation profiles, as well as comparing them across groups with different smoking histories. The results of this experiment, which contributes novel findings to this field of research are summarized in the next section and the appendices.

Third, given the case and control specimens from the original source were suboptimal, we have turned to our second cohort of subjects, a group that has been prospectively developed over the last three years through our Lung Cancer Early Detection and Prevention Clinic. This resource has provided 51 cases of lung cancer thus far. While this is does not meet the original of evaluating epigenetic profiles in 2 independent populations, we intend to continue this work by collecting case and noncase specimens from our clinic. These specimens were recently submitted for genotyping. The data should be ready for analysis within the next month.

All of the specimens were analyzed using the Illumina Methylation Goldengate Cancer panel I, which contains 1,505 CpG loci selected from 807 genes in the following categories: tumor suppressor genes, oncogenes, DNA repair, cell cycle control, differentiation, apoptosis, X-linked, and imprinted genes. In total, we have used the resources from this grant to genotype 167 buccal specimens (52 cases, 116 controls) and 157 blood specimens (50 cases, 107 controls).

Key Research Accomplishments

The key findings from this work are summarized in Appendix A. This comes from work that compared blood and buccal DNA from current, former and never smokers. These findings can be summarized as follows:

- Buccal and blood methylation profiles are highly reproducible. Correlation coefficients for technical replicates of buccal and blood specimens were 0.97 $\pm$ 0.03 and 0.99 $\pm$ 0.005 respectively.
- The methylation patterns of the blood and buccal DNA appeared to be distinct; unsupervised clustering correctly classified the blood and buccal specimens (Appendix B Figures 1-3).
- The epigenetic profile of buccal and blood DNA is most similar among never smokers. The epigenetic profile of buccal and blood DNA is most different among current smokers. This suggests that tobacco smoke exposure does affect buccal DNA differently from blood DNA.
- The methylation profiles of the buccal DNA from current and former smokers were most similar, but very different from never smokers. Smoking exposure was associated with more hypermethylated loci when compared to former smokers.
- Tobacco smoke exposure, either as a current or former smoker, is associated with more hypermethylated buccal DNA than never smokers, suggesting that tobacco smoke exposure may have a biologic effect on the epigenetic profile detected in DNA from the blood, resulting in more methylation differences when compared to never smokers.

This manuscript is currently a work in progress awaiting final data analysis. Thus the draft manuscript included in the appendices represents an early version of the work. We anticipate this will be completed within the next month and be ready for submission for publication.

Reportable outcomes

The findings summarized in the Key Research Accomplishment Section will be reportable outcomes that will be submitted for publication within one month.

Conclusions

This work has confirmed that tobacco smoke exposure does affect the methylation profile of buccal DNA, and that this effect does not extend into blood DNA. These results will be used to guide the next phase of analysis, where we will use the methylation markers most susceptible to tobacco exposure as biomarkers and compare their profiles between lung cancer cases and controls.

References

None

List of Personnel

Jason W. Chien, M.D., M.S.
Wenhong Fan, Ph.D.
Michelle Olson

Appendices

APPENDIX A

Tobacco smoke exposure effects on the epigenetic profiles of buccal and blood DNA

Jason W. Chien [1,3], Wen Hong Fan [2], Michelle Olson [1], Jeffery Stevens [1], Lue Ping Zhao [1]

[1] Clinical Research Division and [2] Public Health Sciences Division at the Fred Hutchinson Cancer Research Center, Seattle WA, and [3] Division of Pulmonary and Critical Care Medicine, University of Washington, Seattle, WA

DRAFT

ABSTRACT

INTRODUCTION

METHODS

Patient population

Our Institutional Review Board has approved all of the activities conducted in this study. Patients who were clinically evaluated in the Lung Cancer Early Detection and Prevention Clinic (LCEDPC) at the Seattle Cancer Care Alliance between January 1, 2008 and July 30, 2010 were eligible for this study. This clinic evaluates two main patient populations: 1) individuals at high risk for lung cancer who would like to have their lung cancer risk assessed, and 2) patients with intra-thoracic lesions (e.g. pulmonary nodules, mediastinal adenopathy, lung masses, endobronchial lesions) that are suspicious for lung cancer. All patients were evaluated according to standard clinical protocols. All participants diagnosed with a cancer of any type or with a history of cancer were excluded from participating in this study. Participants included in this study were confirmed to have no lung cancer if their chest computed tomography scans met one of three criteria: 1) there is no evidence of a pulmonary lesion suspicious for cancer, 2) pathologic confirmation that a pulmonary lesion is not lung cancer or 3) there was no interval growth of a suspicious pulmonary lesion over a minimum of two years.

Clinical data

Standard demographic data were collected for all patients (see Table 1). A detailed smoking history was obtained by designating whether each participant was a never smoker (defined as <100 cigarettes smoked during lifetime), former smoker (quit for at least one year) or current smoker. Smoking histories were determined by identifying the age that smoking started, the average number of cigarettes smoked per day, and if applicable, when the individual quit smoking. Smoking exposure was then calculated as pack-years by multiplying the number of total years smoked by the number of packs of cigarettes smoked each day.

Biologic specimens

Blood and buccal specimens were collected from each participant at the end of their first clinic visit. Under direct supervision from the study coordinator, each participants was required to rinse their mouth with tap water for ten seconds, then brush their inner cheek with cytology brushes to collect the buccal cells. The head of each cytology brush was then cut off and placed in a 1.5ml microcentrifuge tube and stored at -80ºC until DNA isolation. DNA was isolated from whole

blood and buccal brushes using the QIAmp DNA mini kit (Qiagen, Valencia, CA) according to manufacturer directions, and then subjected to bisulfite conversion using the EZ-96 DNA methylation kit (Zymo Research, Orange, CA). DNA quality and quantity were measured using a Nanodrop 1000 Spectrophotometer.

Methylation analysis

The bisulfite converted DNA were analyzed using the Illumina Methylation Goldengate Cancer panel I, which interrogates 1,505 CpG loci selected from 807 genes in the following categories: tumor suppressor genes, oncogenes, DNA repair, cell cycle control, differentiation, apoptosis, X-linked, and imprinted genes (231 genes contain one CpG site pergene, 463 genes contain two CpG sites, and 114 genes have 3 or more CPG sites). Data were reported as beta values, which represent ratios of competitive primers for the unmethylated and methylated bisulfite sequences. Methylation status of the interrogated CpG site was determined using the BeadStudio Software (Illumina, Inc. San Diego, CA), calculated as the ratio of signal from a methylated probe relative to the sum of both methylated and unmethylated probes, and is represented by the value “beta”, which ranges continuously from 0 (unmethylated) to 1 (fully methylated). Detection p-values were computed from the background model characterizing the chance that the target signal is distinguishable from negative controls. Loci with a detection p-value >0.05 were identified as “failures,” with the corresponding beta values deemed missing and removed from the analysis. Because the beta value is dependent upon the number of strands of DNA in the sample, biased beta value are generated for CpG sites on the X chromosome (there are no probes on the Y chromosome). Therefore, all probes on the sex chromosomes were also removed from the analysis.

Data analysis

All analyses were performed using Matlab v.2010b (The MathWorks, Inc. Natick MA, USA). Pearson's r, a measure of the strength of the linear relationship between two variables, was used to evaluate duplicated samples. Paired and unpaired beta values were analyzed using the Student’s t-test. The resultant z-score was plotted in histogram plots to evaluate the extent of epigenetic profile differences between comparison groups. To control false positive error rates, we used the Number of False Discovery (NFD) method, so that the total number of false discoveries, from a list of discoveries, is controlled at a fixed preset number [15]. Conceptually, NFD is a count of the false positive signals, and is closely linked with Bonferroni’s correction, except that interpretations of their numerical values are quite different. NFD is also closely connected with the false discovery rate (FDR). When locus-specific p-values are estimated, and if all tests were independent, NFD equals the number of tests multiplied by p-values (NFD=m*P) [15].

RESULTS

We evaluated paired blood and buccal DNA from 33 current smokers, 22 former smokers, and 16 never smokers. The demographic characteristics of these participants are summarized in Table 1. From the original 1,505 CpG loci, we removed XX failed loci and XX loci located on the X-chromosome, resulting in a total of 1421 loci that were used in these analyses.

As a first step, we assessed assay reproducibility by analyzing the data from technical and plate-to-plate replicates. Technical replicates were performed for 16 buccal and 17 blood samples. The mean $\pm$ standard deviation of the correlation coefficients were 0.97 $\pm$ 0.03 and 0.99 $\pm$ 0.005 respectively. Plate to plate replicates were performed for 33 buccal and 4 blood samples, with mean correlation coefficients of 0.96 $\pm$ 0.03 and 0.99 $\pm$ 0.002 respectively. Given these correlation coefficients reflect that the methylation data was highly reproducible, we did not perform replicates for the remainder of the samples. Because replicate data were available for some specimens, the average beta value of each locus was used in the analyses below (Wenhong, is this correct?).

Comparison of buccal and blood methylation profiles

Differences in epigenetic signatures between buccal and blood DNA were first assessed by using nonsupervised hierarchical clustering to evaluate the beta values for each locus, stratified according to smoking status. For each of the smoking groups, the methylation patterns of the blood and buccal DNA appeared to be distinct; unsupervised clustering correctly classified the blood and buccal specimens (Supplemental Figures 1a-c). To further quantify the magnitude of these differences, we compared the beta values for each locus. Figure 1 provides histogram curves of the z-scores from comparing beta value of each locus as assayed using the buccal and blood DNA among never, former and current smokers. As reflected by the narrower morphology of the never smoker curve, the epigenetic profile of buccal and blood DNA is most similar among never smokers. The broader morphology of the current smoker curve suggest that the epigenetic profile of buccal and blood DNA is most different among current smokers. Unlike the curves for former and never smokers, the histogram curve for current smokers is also shifted slightly towards the positive end of the y-axis, which suggests that overall, these loci were more hypermethylated in the buccal DNA when compared to the blood DNA.

To further quantify these differences, we arbitrarily set a z-score threshold of --10 to +10 (Wenhong, I think the 0.5 threshold is giving us too many loci, making this seem insignficant. Can you use a +/-10 threshold instead) and determined that among never, former and current smokers, there were a total of xx, xx, and xx loci whose z-scores fell outside of this range respectively. Among these, xx loci common to all three groups had z-scores outside this range,

suggesting that the methylation status of these common loci consistently did not correlate between blood and buccal specimens, regardless of smoking status. Conversely, among the loci whose z-scores fell within the -10 to +10 range, xx loci were common to all smoking groups.

Effect of smoking exposure on blood and buccal epigenetic profiles

To evaluate the effect of tobacco smoke exposure on blood DNA, we compared, the blood methylation profiles among the different smoking groups: former versus never, current versus never, and current versus former. The z-score histogram curve for each comparison group was centered over zero on the x-axis, although the z-score spectrum was narrower than observed in the previous comparison (Figure 2). The blood methylation profile for the current versus former smokers was most similar, only xx loci had z-scores outside the -2 to +2 range (a tighter range was used due to the overall narrower distribution of the z-scores) (Wenhong, again, I think this range results in fewer loci outside the range, which makes the result seem a little more credible). The blood methylation profile was more different when current and former smokers were compared to never smokers. For the former versus never and current versus never smokers, xx and xx loci had z-scores outside the -2 to +2 range, respectively.

This same analysis was performed using buccal DNA. The buccal z-score histogram curves were visibly different from the blood DNA curves (Figure 3). As again demonstrated by the narrow morphology of the current versus former smoker curve, the methylation profiles of the buccal DNA from current and former smokers were most similar; xx loci had z-scores outside the -2 to +2 range. However, there was a shift of the current versus former curve toward the positive end of the z-score spectrum, indicating that smoking exposure was associated with more hypermethylated loci when compared to former smokers. Indeed, of the xx loci with z-scores outside the -2 to +2 range, xx loci had z-scores > +2. While the current versus never smoker curve is broader than the current versus former smoker curve, this curve is also shifted to the positive end of the y-axis; xx loci had z-scores outside the -2 to +2 range, with xx of these > +2. The curve for former versus never smokers was broader than the current versus never smoker curve, but it is centered over zero on the y-axis; xx loci had z-scores outside the -2 to +2 range. These observations indicate that tobacco smoke exposure, either as a current or former smoker, is associated with more hypermethylated buccal DNA than never smokers, suggesting that tobacco smoke exposure may have a biologic effect on the epigenetic profile detected in DNA from the blood, resulting in more methylation differences when compared to never smokers.

Application of unsupervised hierarchical clustering to all 1421 loci was not able to reliably group the blood or buccal specimens according to smoking groups (Supplemental Figure 2a and 2b). However, we repeated the analysis using a subset of loci whose z-scores were >+2 or < -2 and were common to all of the comparison groups. There were xx and xx loci that met these criteria for blood and buccal DNA respectively (Table 2, Wenhong, need this table).

Unsupervised hierarchical clustering using these loci revealed that the loci identified in the buccal DNA analysis was able to reliably differentiate between current, former, and never smokers (Figure 4a, Wenhong need this figure). However, these groups was not able to be differentiated using the loci identified in the blood DNA analysis (Figure 4b, Wenhong, need this figure) NOTE, SINCE I HAVE NOT SEEN THE RESULTS, THESE STATEMENTS ARE JUST GUESSES.

Table 1. Participant characteristics

Characteristic	Never smoker	Former smoker	Current smoker
Number of participants	16	22	33
Median age (IQR)	59 (±12.5)	60.5 (±21)	54 (±15)
Sex			
Male (%)	4 (25)	12 (54.5)	18 (54.5)
Female (%)	12 (75)	10 (45.5)	15 (45.5)
Race			
White (%)	11(68.8)	18(81.8)	29(87.9)
Non-white (%)	5 (31.2)	4 (18.2)	4 (12.1)
Median pack years (IQR)	NA	22.5 (±18)	26 (±24.5)
Median years quit	NA	19.5 (±18)	NA

APPENDIX B

Figure 1

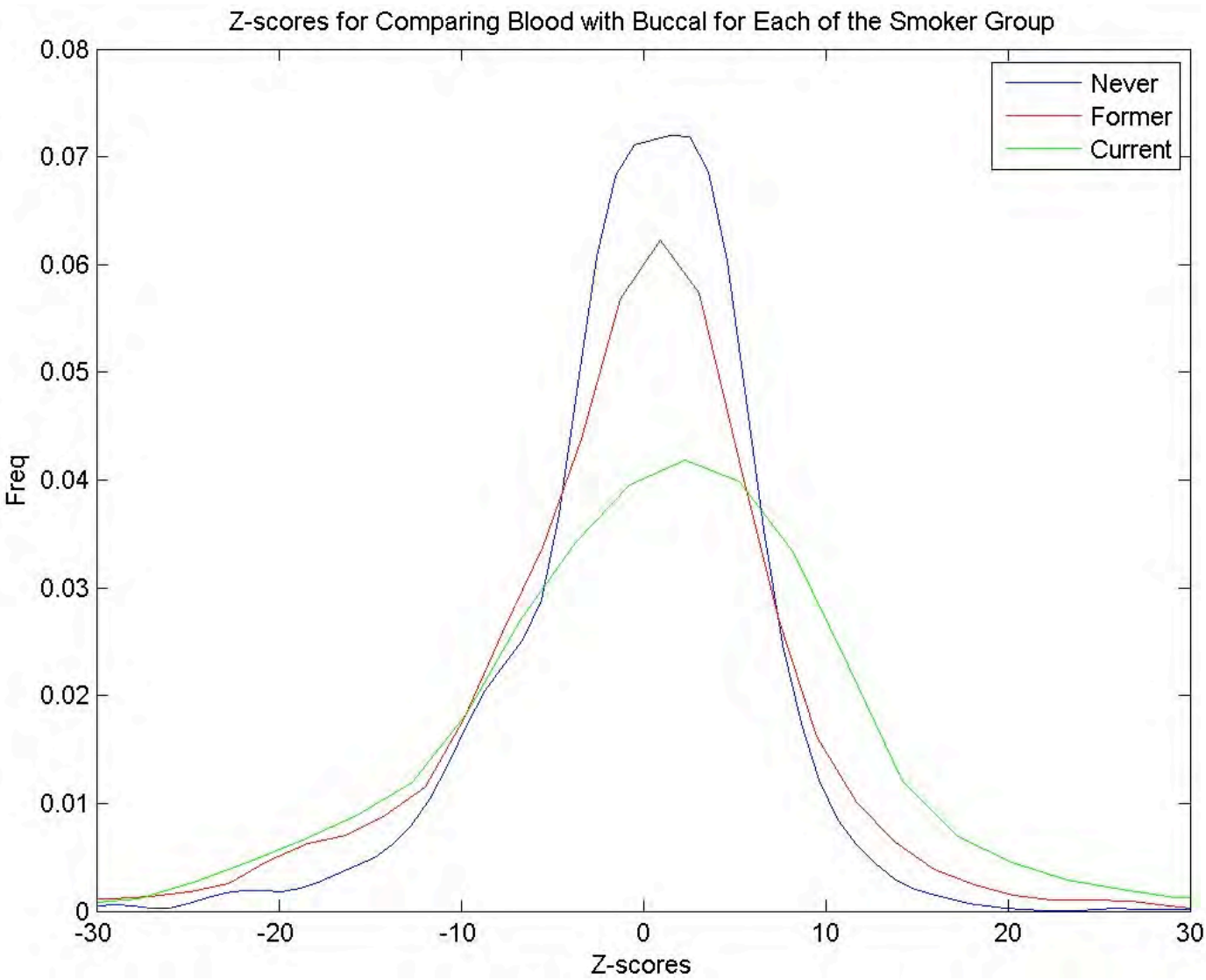

Z-scores for Comparing Blood with Buccal for Each of the Smoker Group
Never
Former
Current
Freq
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
-30
-20
-10
0
10
20
30
Z-scores

Figure 2

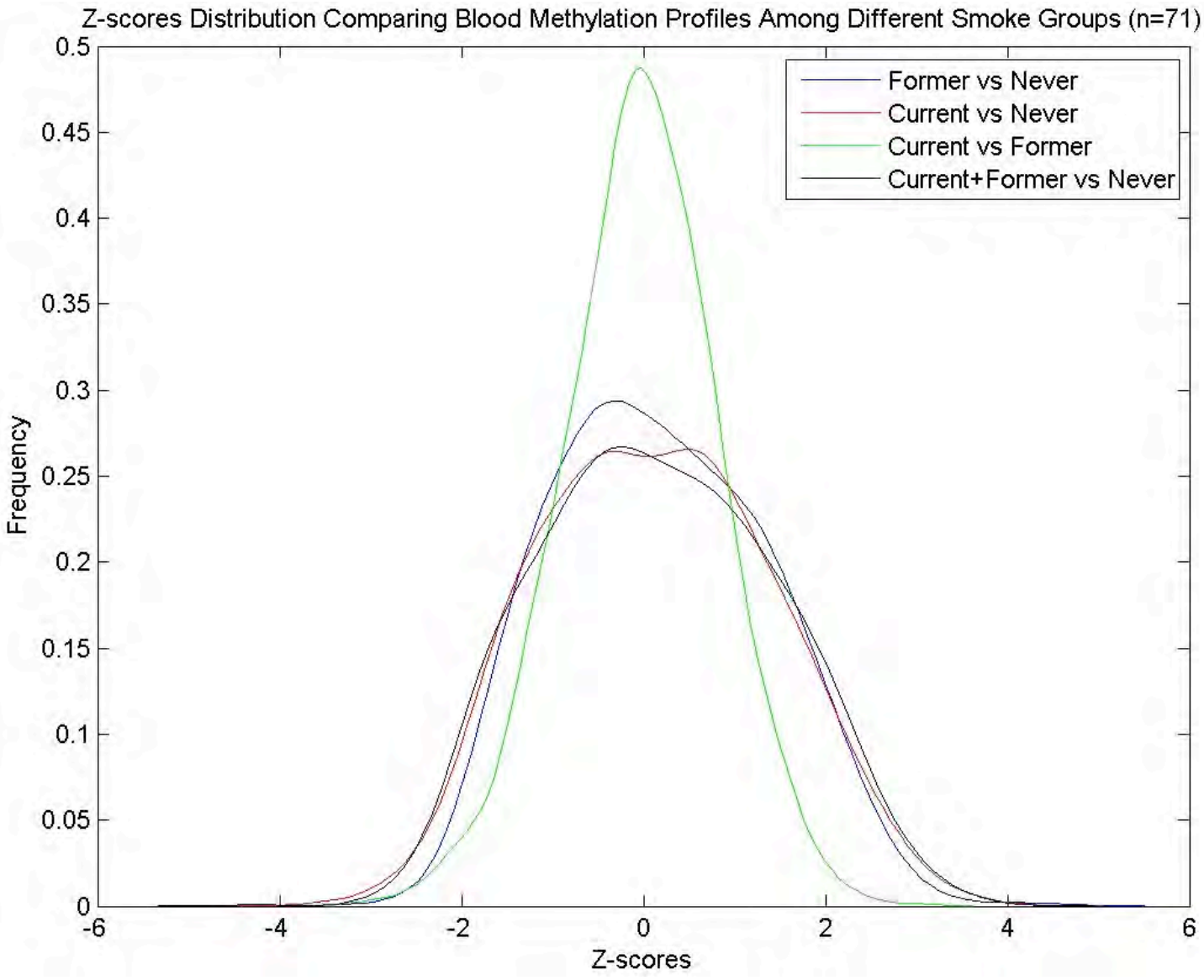
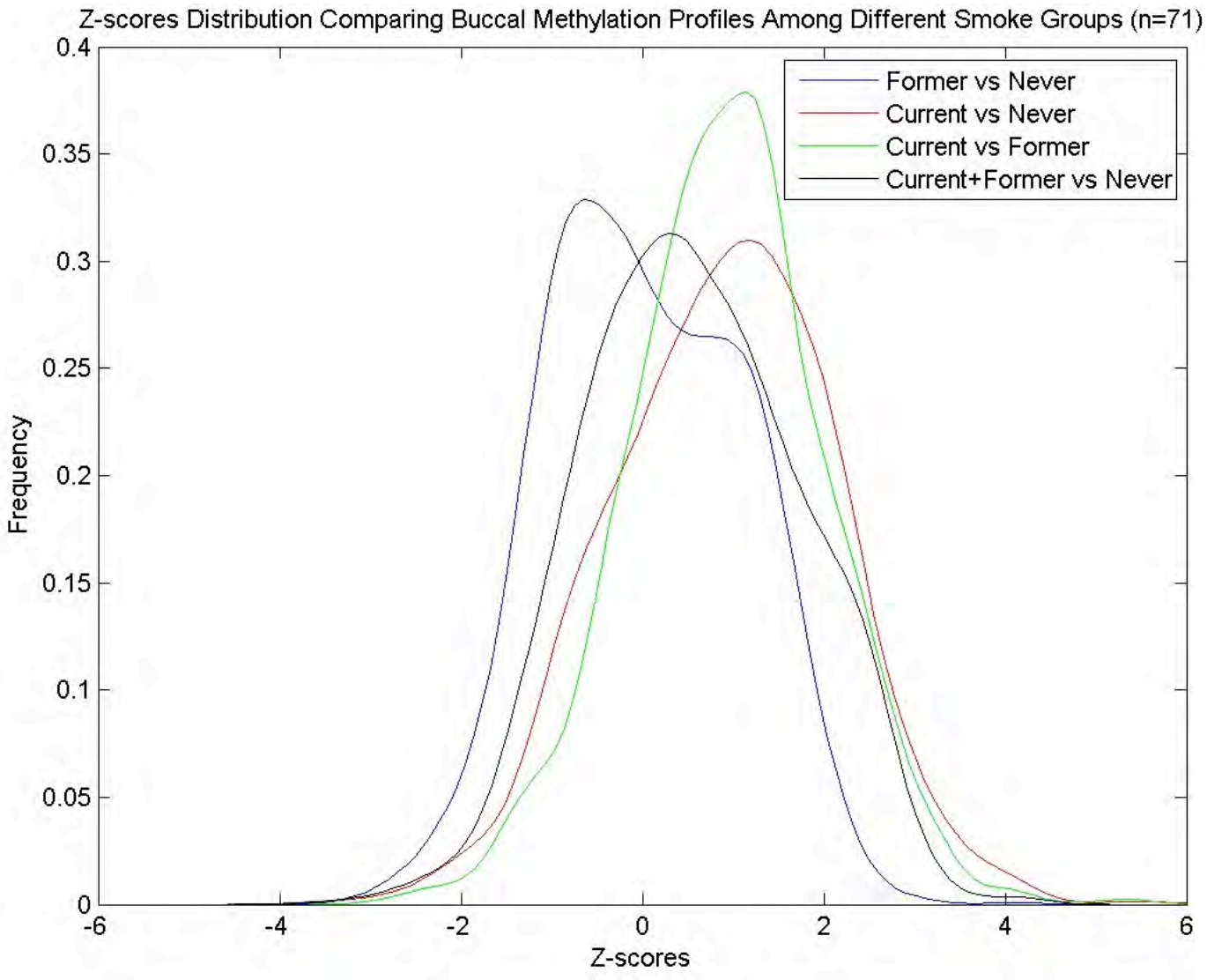

Z-scores Distribution Comparing Blood Methylation Profiles Among Different Smoke Groups (n=71)
Former vs Never
Current vs Never
Current vs Former
Current+Former vs Never
Frequency
Z-scores
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-6
-4
-2
0
2
4
6

Figure 3


Z-scores Distribution Comparing Buccal Methylation Profiles Among Different Smoke Groups (n=71)
Former vs Never
Current vs Never
Current vs Former
Current+Former vs Never
Frequency
Z-scores
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-6
-4
-2
0
2
4
6